2

Manpower R&D Program

Grant #N00014-91-J-1361

Development of Intelligent, Computerized Aids for the Specification of Causal Models From Large Data Bases

Richard Scheines, Peter Spirtes, and Clark Glymour

Final Report

May 25th, 1992

1. Introduction and Summary

In this and two previous grant periods, our research group has developed TETRAD II and used it in collaboration with researchers at NPRDC to analyze data sets pertinent to manpower issues, e.g. recruiter satisfaction (Scheines 90). TETRAD II is a suite of tools to aid in the discovery of causal models of statistical data. The work during this grant period focussed on three areas: 1) making TETRAD II friendly enough to be used productively by non-developers, 2) developing a new module of TETRAD II that helps discover causal structure among latent, or unmeasured, variables, and 3) to use this tool to help analyze the NPRDC Youth Attitude Tracking Survey (YATS) data set.

To make TETRAD II friendly enough for non-developers, we stabilized and debugged the program, and wrote and tested a 200 page users manual, which is available upon request. In October 1991, we installed the program at NPRDC and distributed the user's manual to several researchers there.

The module that helps discover causal structure among latent variables, MIMBuild,[1] has been completed and is a part of the TETRAD II program delivered to NPRDC in October 1991. It is described in two papers written during the grant period, (Scheines 91 and Scheines 92). The module has been tested on simulated data (Scheines 92), on real data (Callahan 92) and its algorithm has been proved correct (Spirtes 92).

Stephen Sorensen of NPRDC and Jan Callahan of Callahan Associates have used an early version of MIMBuild to construct a latent variable model of civilian attitudes toward a military career. They found that as a tool, MIMBuild compares favorably to more standard factor analytic techniques. The report their results in "Using TETRAD II as an Automated Exploratory Tool" a paper which is due to appear in the 1992 fall issue of Sociological Methods and Research, which received a Certificate of Commendation from The Chief of Naval Research in recognition of nomination for Best FY-90 Independent Exploratory Development Project, and which we include in this report as an Appendix.

Carnegie-Mellon Univ., Pittsburgh, PA

per telecon ONR 6/17/92 cl

[1]MIMBuild stands for the Multiple Indicator Model Builder.

2. TETRAD II

TETRAD II now has seven modules, each of which has its own chapter in the user's manual. In summary, they are:

Build - Takes covariance, correlation, or categorical data and outputs a class of causal structures that explain the independence and conditional independence relations judged to hold in the population from which the sample is taken.

Search - Takes covariance data and an initial structural equation model with latent variables and outputs a class of respecifications of the model.

Mimbuild - the **M**ultiple **I**ndicator **M**odel **Build**er. Takes a list of latent variables, a set of indicators for each latent variable, and the covariance or correlation matrix among all of the indicators, and outputs a class of causal structures among the latent variables.

Monte - the Monte Carlo Generator. Takes a causal structure and information about its interpretation, and outputs sample data generated pseudo-randomly from the distribution characterized by the interpreted and parameterized causal structure.

Estimate- Takes a Bayesian network and categorical data, and provides a maximum likelihood estimate of the joint distribution that satisfies the independence constraints imposed by the Bayesian network's causal structure.

Update - Takes a parameterized Bayesian network and values for a subset of its variables, and produces the conditional distribution of the remaining variables.

EQSwriter - Takes a causal structure, interpreted as a linear structural equation model, and covariance data, and produces an input file to EQS, a popular estimation package for structural equation models.

The program and its source code are installed on a Sun 4 Workstation at NPRDC supervised by Stephen Sorensen. We have already updated the program twice, and will continue to do so periodically. Besides including detailed instructions on each of TETRAD II's modules, the user's manual also includes an introduction explaining the basic theoretical ideas that the program is based on (chapter 1), a guide to the research contexts appropriate for each module (chapter 1), and a series of applications to illustrate the use of each module (chapter 11).

In the future we hope to improve the program's interface by shifting from a command based interface to one that is menu driven.

3. MIMBuild

3.1 The Problem Addressed

MIMBuild helps specify multiple indicator models. A **multiple indicator model** is a structural equation model with latent variables in which each latent variable has at least two measured indicators. Multiple indicator models are typically divided into a "structural model," which is a system of simultaneous equations among latent factors (equation 1), and a "measurement model," which is a system of equations describing the relations between latent and measured variables (equation 2). The structural model can be given by[2]

[1] $$\eta = B\eta + \zeta$$

where η is a (m x 1) vector of latent factors, ζ is a (m x 1) vector of disturbances, B is a (m x m) matrix of coefficients, and $Var(\zeta) = \Psi$. The measurement model can be given by

[2] $$y = \Lambda\eta + \varepsilon$$

where y is a (p x 1) vector of observed variables, Λ is a (p x m) matrix of factor loadings, ε is a (p x 1) vector of disturbances, and $Var(\varepsilon) = \Theta_\varepsilon$.

Such models can be parameterized by a vector θ of the exogenous variances (and covariances) in Ψ and Θ_ε, and the linear coefficients in B and Λ. Specifying a particular θ produces a covariance matrix $\Sigma = \Sigma(\theta)$ among the observable variables y. Various techniques exist for estimating $\hat{\theta}$, e.g., full-information maximum likelihood (ML) as implemented in LISREL or EQS. Our concern, however, is with specification, not estimation.

[2]We neglect the usual distinction between exogenous (ξ) and endogenous (η) latent variables. Any model can be written solely in terms of η variables (Bollen, 1989), and the standard LISREL formalism for expressing structural relations is much more flexible for η variables than it is for ξ variables.

Let the **causal structure** of a model be given by η, y, and those elements of B, Λ, Ψ, and Θ_ε that are *not* fixed at 0. A causal structure can be represented without any loss of generality by a diagram in which only those elements of B, Λ, Ψ, and Θ_ε that are not fixed at 0 are pictured. From the diagram in figure 1, for example, we can infer that Θ_ε is diagonal.

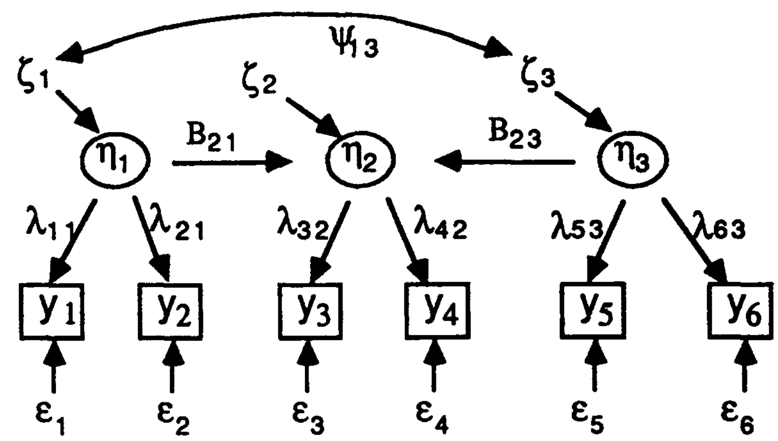

Figure 1

Besides giving a causal structure, specifying a model can involve imposing constraints on θ. For example, certain parameters can be fixed at non-zero values, sets of parameters can be left free but constrained to be equal, etc. Here we will not consider models with these types of constraints.

The problem with model specification is that often the space of plausible alternatives is vast, even for relatively small models in which a lot of background knowledge is available. The following example illustrates this point.

In his book, *Structural Equation Models with Latent Variables*, Ken Bollen discusses an example from social psychology analyzed by Marsh and Hecevar (1985). The model involves four latent variables concerning self-concept for fifth graders. Bollen's example employs the measurement model in figure 2, where the latent variables are each self-measures of the following four dimensions:

η_1 = Self-Image of Physical ability
η_2 = Self-Image of Physical appearance
η_3 = Self-Image of Relations with peers
η_4 = Self-Image of Relations with parents

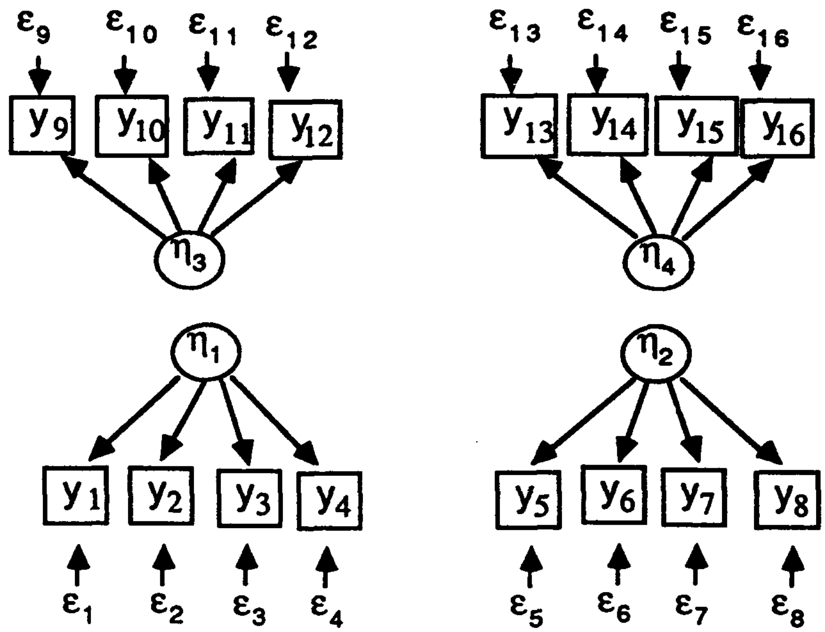

Figure 2

The example illustrates how difficult it can be to locate the model or class of models that best fit the data. Suppose that we were quite confident that the measurement model as specified was correct. That is, suppose we believed that no two error terms are correlated, that no error term is correlated with a latent variable, and that each indicator measures the latent it is connected to and no other. Although in this case we have no reason to believe so, suppose for illustrative purposes that we were also willing to simplify matters by assuming that the latents we specified are *causally sufficient*, i.e., that no pair of the η variables are effects of a latent not already included in our model.[3] Then all that remains is to specify the strucutral model among η_1 - η_4. To do this, we have to specify, for each pair $<\eta_i,\eta_j>$, whether η_i causes η_j, η_j causes η_i, or whether there is no direct relation between the two.[4] Since all of the latent variables are self measures, there is little reason to prohibit causal connections between them *a priori*. Thus even with only four ηs, there are still 729 different structural models, most of which are identified. If there were 12 ηs, and we were willing to make the extra assumption that the model was recursive, then the number is truly astronomical (Harary 73): 521,939,651,343,829,405,020,504,063.

[3]In contrast Bollen uses this example to illustrate a second-order factor model.

[4]We do not consider models in which latent variables can influence each other, i.e., simultaneous equation models.

In general, the number of models among a set of n variables is the number of ways each pair of variables can be connected, raised to the power of the number of pairs. For causally sufficient structures this is $3^{\binom{n}{2}}$, which is on the order of 3^{n^2}. Although it may be possible to use theory to reduce the number of alternative models (at the very least it must be possible to use theory to eliminate all unidentified models), typically a large number of eligible models will still be left. For structures that are not assumed to be causally sufficient, the numbers are even worse.

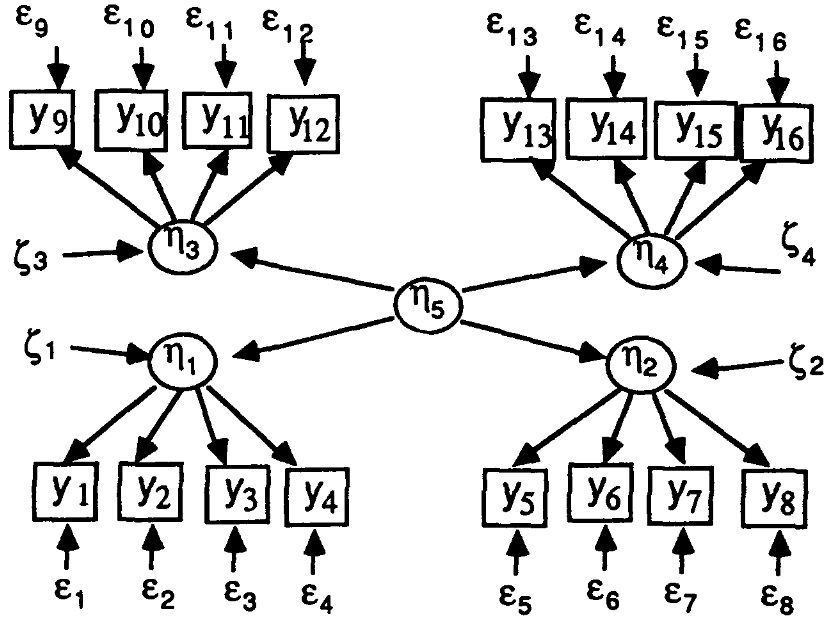

Figure 3

Bollen in fact does not assume that the four latent variables above are causally sufficient. Although his purpose is only illustrative, as is ours, he includes a second-order factor η_5, "general non-academic self-image," that is a common cause of all four latent variables above. This model (figure 3) fails a chi-square test with $p < .001$. Bollen does not attempt to respecify the model so that it fits the data better, but Marsh and Hecevar estimate a host of models with 2nd and 3rd order factors, and find no model with $p(X^2) > .001$. So even though in this case the researchers were quite confident of their measurement model, which accounts for the bulk of the specification and is highly overidentified, they were unable to find any model which was not rejected by a weak significance test. Moreover, because they were not able to perform a very exhaustive search, they could not be confident that there were not other overidentified models that fit the data better than the ones they considered.

The general point is this: even if we are quite confident of large parts of our model, but unsure of the specification of a small portion, the number of alternatives may still be too big to investigate thoroughly. Even if we manage to specify a few models that pass a statistical test, how can we accept the conclusions they suggest unless we know that all alternatives over which we are theoretically indifferent but which lead to different conclusions fail the same test? What is needed is a computational procedure that

1) can identify the smallest class of models that are consistent with background knowledge and that fit the data best, and
2) is fast enough to be practical on models of realistic size.

3.2 Starting Points

No computational aid to specification can replace the scientist. Someone must choose the set of variables to measure, they must identify and interpret the latent variables, and they must insure that the models specified are identified and cohere with established knowledge. Different scientists attempting to build a latent variable model begin the specification task with different sorts of knowledge. In panel models time order and symmetry can play a large role, while in some studies the data are entirely cross-sectional. Background knowledge might rule out certain connections, require others, make some connections probable, and be totally indifferent over the rest. This sort of background knowledge might be representable within a Bayesian framework, and some researchers have begun to explore this possibility (Cooper 91, Buntine 91). Incorporating background knowledge optimally is a topic that we hope to address systematically in future research. For present purposes, we asked ourselves the following question: what parts of a model's specification is a scientist often confident about? Our answer: an overconstrained measurement model.

In many studies, especially those that involve survey data, the relevant set of latent variables and measures for them can be specified with some confidence. That is, in many studies a measurement model can at least be partially specified. What is usually in doubt about such measurement models is not the connections asserted to exist, but whether there are other unspecified connections that might also exist. For example, if subjects are asked the question: "Do you like your job?" in 1975 and again in 1980, no one doubts that their answers measure the latent variables "Job Satisfaction in 1975" and "Job Satisfaction in 1980." Few would be confident *a priori* that the disturbances of these questions are uncorrelated, however. In many cases it is also unclear whether an indicator loads only on the latent variable it was intended to measure.

We have therefore chosen to assume that the modeller has chosen an appropriate set of latent variables, and has chosen at least two indicators for each latent variable that actually measure the latent variable. We also assume that the modeller has no additional information that would help him or her distinguish among the set of possible identified models. Although this assumption is obviously never true in practice, it represents in some sense the worst case. Any knowledge that can be brought to bear to narrow the class of structural models, or order them in some way, can only help matters. Incorporating such knowledge explicitly into the procedures we discuss below is a research topic we hope to address in the future.

We impose three conditions on the initially specified measurement model $y = \Lambda\eta + \varepsilon$:

1) no row in Λ has more than one non-zero,
2) no column in Λ has less than two non-zeros, and
3) Θ_ε is diagonal.

We also impose an additional assumption about the relationship between the initially specified model and the true measurement model, namely:

4) if $\Lambda(i,j) \neq 0$ in the specified model, then $\Lambda(i,j) \neq 0$ in the true model.

The first three conditions describe features of the initially specified model. The first says that every indicator loads on only one latent variable, the second says that each latent variable is measured by at least two indicators, and the third says there are no correlated disturbances. Gerbing and Anderson (1982) call a measurement model of this sort *uni-dimensional*. Scheines (1992) calls it a *pure* measurement model. The last condition indicates that in the true model, any indicator that is initially specified to measure a latent variable does in fact measure the latent variable. Such a measurement model might still be misspecified in three ways:

1) Indicators might measure more than one of the specified latent variables,
2) Error terms might be correlated, and
3) Indicators might be directly related.

Correcting these sorts of specification errors, if we could identify them, involves only freeing parameters initially fixed at zero.[5] Thus we say that the initially specified measurement model is *strictly overconstrained.*

3.3 Tetrad Differences and Multiple Indicator Models

Tetrad differences involving different foursomes of indicators provide different types of specification tests, including tests for the purity of the measurement model. These tests have three main advantages over Langrange Multiplier types of tests, which are used by LISREL, EQS, and CALIS. First, they are localized tests in that they focus on one part of the model and do not require a specification of the other parts of the model. As such, their reliability does not depend on the full structure being correctly specified, which is exactly what cannot be guaranteed when the true model is unknown. Second, they test a number of overidentifying restrictions simultaneously and thus can economize on the number of separate tests that have to be performed. Third, they are analytic tests and thus can be computed quickly. The last two advantages mean that searches based on tetrad differences can be much more exhaustive than say Lagrange Multiplier type searches. The first advantage speaks to the reliability of searches based on tetrad differences.

To illustrate, consider how tetrad differences can be used to isolate pure indicators in the measurement part of the model. Suppose we want to test whether four indicators y1-y4 of a single latent variable are pure. This is equivalent to testing whether the disturbances ε_1-ε_4 in figure 4 are uncorrelated, as in Figure 4-a. Tetrad differences among the four indicators vanish only if no pair of the disturbances are correlated (Spearman 04, Gerbing and Anderson 82, Glymour, et.al, 87). Thus, tetrad differences provide an analytical way of simultaneously testing whether all the disturbances are uncorrelated without having to specify the rest of the model, including the structural model.

[5] In fact correcting the third sort of specification error is more complicated. It involves rewriting the model so that each indicator is actually represented as an η variable so that it is possible to represent the possibility that indicators directly influence each other. See (Bollen 89).

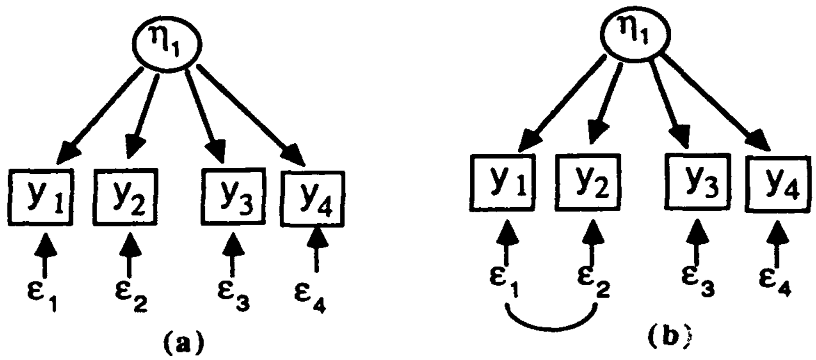

Figure 4

Now consider how tetrad differences can be used to test whether the disturbances in the indicators of one latent variable are uncorrelated with the disturbances of the indicators of another latent variable. Tetrad differences among two indicators y1,y2 of one latent variable $\eta 1$ and two indicators y3,y4 of another latent variable $\eta 2$ vanish (figure 5-a) only if no cross-construct pair of error terms are correlated (figure 5-b), *regardless of the type of connection between* $\eta 1$ *and* $\eta 2$ (Gerbing and Anderson 82, Glymour et.al., 87). Once again tetrad differences provide an analytical way of testing a set of constraints without having to specify the full model.

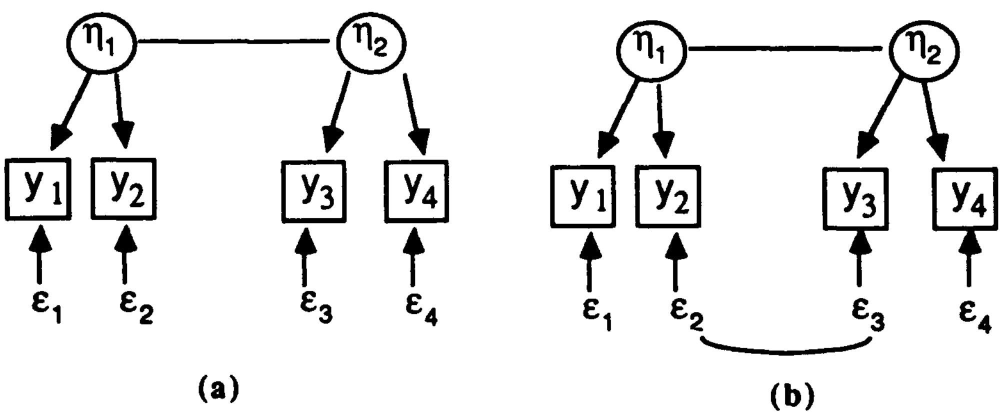

Figure 5

Tetrad differences can also be used to test whether indicators load on only one latent variable. For example, tetrad differences among three indicators y1,y2,y3 of a latent variable $\eta 1$ and one indicator y4 of another latent variable $\eta 2$ vanish (figure 6-a) regardless of the type of connection between $\eta 1$ and $\eta 2$ only if 1) no pair of disturbances ε_i-ε_j are correlated (figure 6-c), and no indicator of $\eta 1$ loads on $\eta 2$ (figure 6-b) (Glymour et.al., 87).

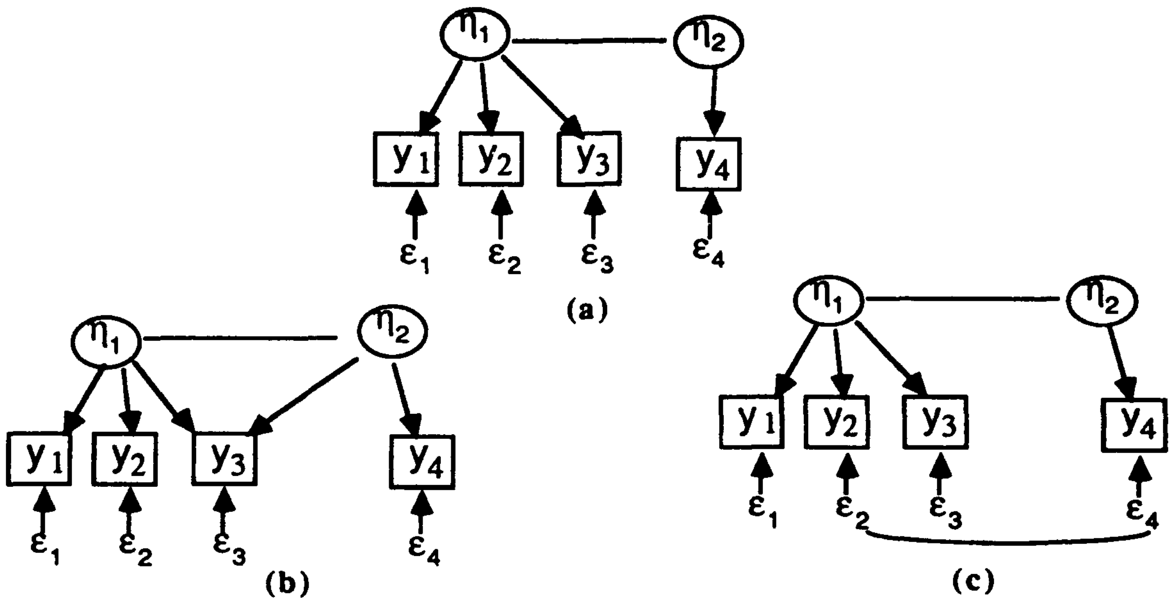

Figure 6

Thus, if each latent variable has at least four indicators in the initial model, then the same type of test can be used to detect impurities in the measurement model *regardless of the structural model.*

A different sort of tetrad difference provides a specification test for certain aspects of the structural model. If the measurement model is pure, then all three possible tetrad differences among two indicators y1,y2 of a latent variable $\eta 1$, one indicator y3 of another latent variable $\eta 2$, and a fourth indicator y4 of a third latent variable $\eta 3$ are strongly implied to vanish (figure 7) only if the model strongly implies that $\rho_{\eta 2, \eta 3.\eta 1} = 0$ (Spirtes 92). If $\rho_{\eta 2, \eta 3.\eta 1} = 0$, then there can be no direct relation relation between $\eta 2$ and $\eta 3$.

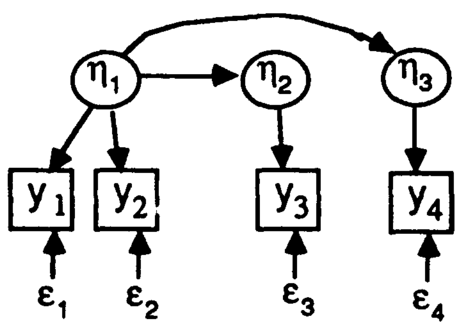

Figure 7

Still another specification test for the structural model is possible if the measurement model is pure. If y_i is an indicator of $\eta 1$ and y_j is an indicator of $\eta 2$, then $\rho_{\eta 1,\eta 2} = 0$ if and only if $\rho_{yi,yj} = 0$.

3.4 The MIMbuild Procedure

MIMbuild takes as input the sample covariances of the observables and a strictly overconstrained measurement model.[6] MIMbuild has been tested on models involving as many as ten latent variables and 50 indicators. Because its complexity is a polynomial function of the number of measured variables instead of an exponential one,[7] it runs on such models in seconds instead of hours or days, and we are confident that it will run in feasible time even on models with over 100 measured variables.

MIMbuild proceeds in three stages. In the first it eliminates impure indicators until it is left with a uni-dimensional, or pure, measurement model. In the second it identifies which pairs of latent variables are adjacent in the structural model,[8] and in the third it orients, or provides a causal direction to, as many of the connections as it can. For example, suppose that the true model is as below (figure 8), and the initial, conjectured model input to MIMbuild is the measurement model in figure 9.

[6] A prototype version of MIMbuild is described in detail in, Scheines, R. and Spirtes, P., (1992) "Finding Latent Variable Models in Large Data Bases," forthcoming in a special issue of the International Journal of Intelligent Systems, ed. Greg Piatetski-Shapiro.

[7] The algorithm's complexity is bounded by the number of tetrad differences it must check, which in turn is bounded by the number of foursomes of measured variables. If there are n measured variables the total number of foursomes is on the order of r^4. We don't check each possible foursome, however, and the actual complexity depends on the number of latent variables and how many variables measure each latent. If there are m latent variables and s measured variables for each, then the number of foursomes is $O(m * s^4)$. Since $m*s = n$, this is $O(n * s^3)$, which is much lower than $O(n^4)$ if $s << n$.

[8] Two variables X and Y are adjacent if either X is a direct cause of Y or Y is a direct cause of X.

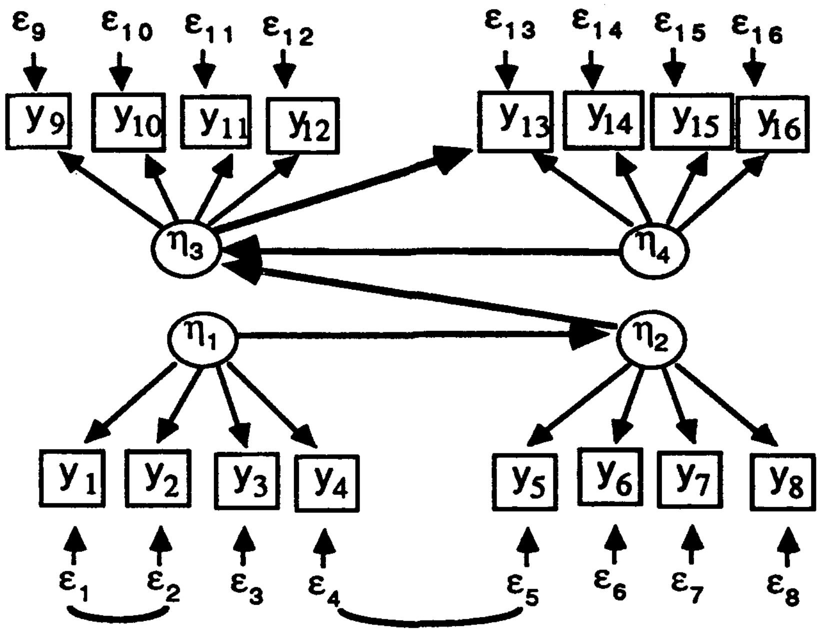

Figure 8: The True Model

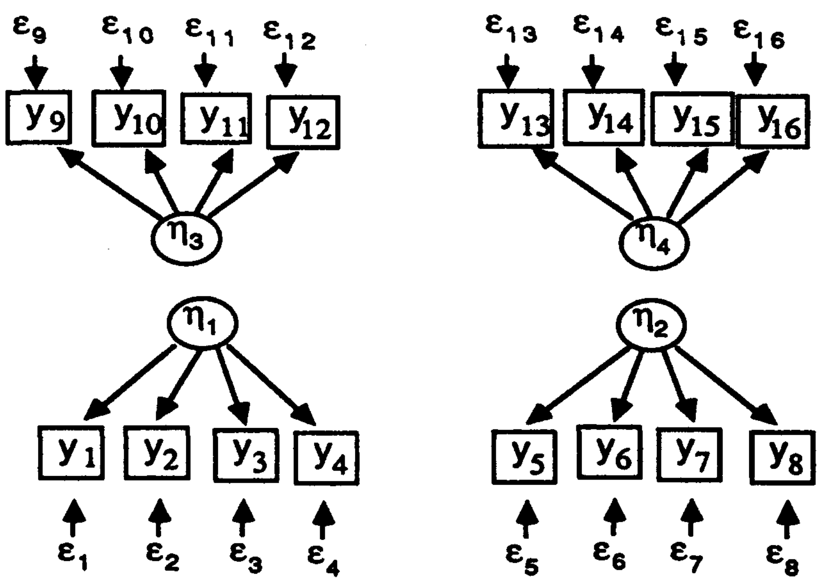

Figure 9: The Measurement Model Input to MIMbuild

In order to deploy the tetrad difference specification test for the structural model, the measurement model must be pure. Thus the first part of the strategy in MIMbuild is to locate and discard those indicators that are impure. Five of the indicators in figure 8 are impure:

y1,y2,y4,y5,and y13. All five do not need to be discarded to achieve a pure measurement model. Removing any of the following sets will do: {y1,y4,y13} {y1,y5,y13} {y2,y4,y13} {y2,y5,y13}. The procedure used in MIMbuild to choose which set of indicators to eliminate is discussed below. Suppose MIMbuild removes the first of these sets of indicators, resulting in the pure measurement model pictured in figure 10.

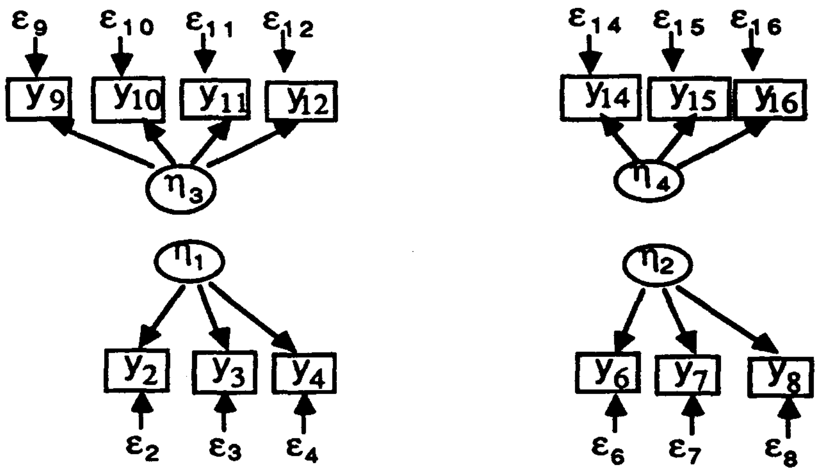

Figure 10

At this point it tests for correlation among each pair of latents by checking for correlation among their indicators. In the above example, it would find that $\rho_{\eta 1,\eta 4} = \rho_{\eta 2,\eta 4} = 0$. Finally, for any pair of latents that it judges are correlated, it checks whether they are correlated when partialed on a third latent. To do this it uses tetrad differences of the final sort described above. In this example, it would find that $\rho_{\eta 1,\eta 3.\eta 2} = 0$. Accordingly, MIMbuild would output the partially directed causal structure in figure 11.[9]

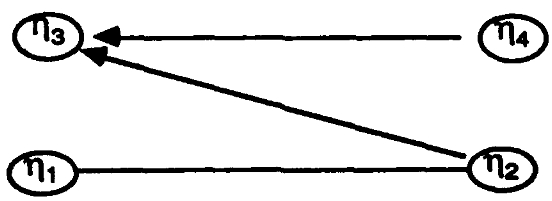

Figure 11

[9]In this diagram and those that follow, for simplicity we have ommitted the disturbance terms. They should be thought of as there implicitly.

The edge between $\eta 1$ and $\eta 2$ is undirected because MIMbuild does not have enough information to orient it. It really means that MIMbuild is outputting the two structures shown in figure 12.

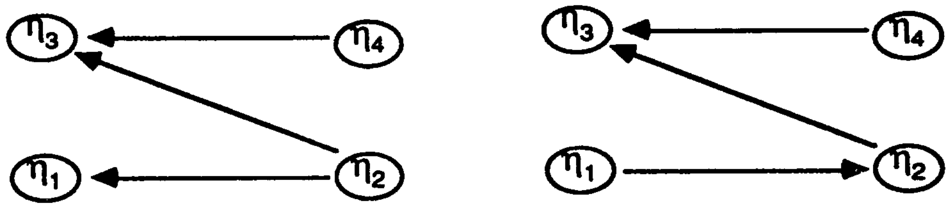

Figure 12

3.5 Simulation Studies of MIMBuild's Reliability

To test the behavior of the procedure on sample data, we used TETRAD II's Monte Carlo generator to produce data from the causal graph in figure 13, which has 11 impure indicators.

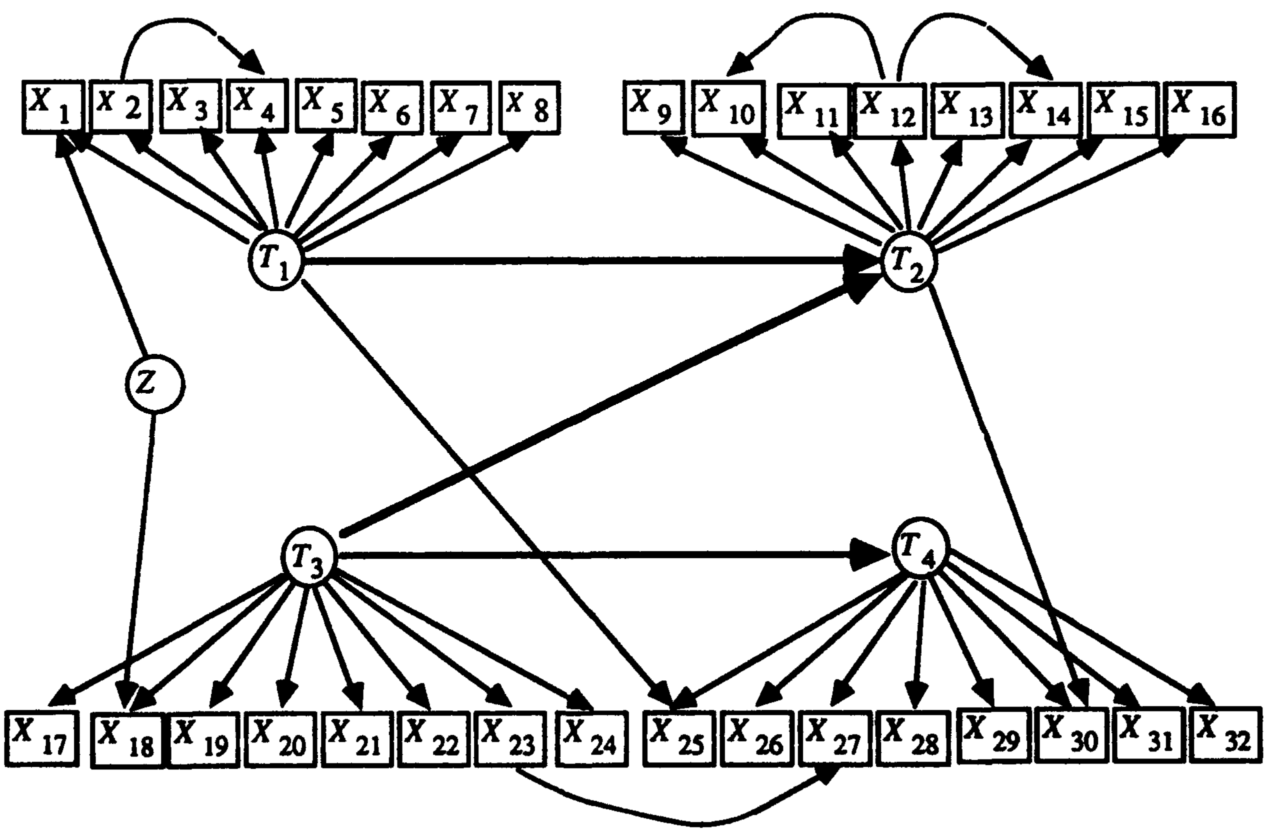

Figure 13

$$\text{Impure Indicators} = \{X_1, X_2, X_4, X_{10}, X_{12}, X_{14}, X_{18}, X_{23}, X_{25}, X_{27}, X_{30}\}$$

The distribution for the exogenous variables is standard normal. For each sample, the linear coefficients were chosen randomly between .5 and 1.5.

We conducted 20 trials each at sample sizes of 100, 500, and 2000. We counted errors of commission and errors of omission for detecting uncorrelated latents (0-order) and for detecting 1st-order d-separation. In each case we counted how many errors the procedure could have made and how many it actually made. We also give the number of samples in which the algorithm identified the d-separations perfectly. The results are shown in the next table, where the proportions in each case indicate the number of errors of a given kind over all samples divided by the number of possible errors of that kind over all samples.

Sample Size	0-order Commission	0-order Omission	1st-Order Commission	1st-Order Omission	Perfect
100	2/80	0/40	7/220	1/20	13/20
500	1/80	0/40	2/220	0	19/20
2000	0	0	0	0	20/20

Extensive simulation tests with a variety of latent topologies for as many as six latent variables, and normally distributed variables, show that for a given sample size the reliability of the procedure is determined by the number of indicators of each latent and the proportion of indicators that are confounded. Increased numbers of pure indicators make decisions about d-separability more reliable, but increased proportions of confounded variables makes identifying the pure indicators more difficult. For large samples with ten indicators per latent the procedure gives good results until more than half of the indicators are confounded.

4. YATS

As Sorensen and Callahan discuss in their appended paper, the Youth Attitude Tracking Survey provides an excellent test case for the MIMBuild procedure. The Survey consists of hundreds of questions that are reasonable indicators of several latent psychological attitudes. In order to model the relations between such attitudes, a measurement model for each attitude must be constructed. If the measurement models constructed are misspecified, then inferences about the relationships between the attitudes based on them are compromised. Thus a crucial step in the modeling process involves forming such measurement models.

Sorensen and Callahan constructed measurement models with two techniques, one an early precursor to MIMBuild that we called SCALES, and the other standard exploratory factor analysis as implemented in SAS. SCALES does not construct measurement models entirely automatically, and offers no guidance whatsoever on how to construct the structural model among latent attitudes. To construct the structural model, Sorensen and Callahan used the Build different module of TETRAD II, then called Partial. The most recent version of MIMBuild automatically constructs the measurement model and then automatically outputs a class of structural models. Even working with a diminished tool-set, however, Sorensen and Callahan were able to locate several full structural equation models for the civilian group that fit the data quite well ($p(X^2) > .7$) and were intuitively quite plausible. Given that the sample size of their data set exceeded 7,000, and that they were not involved in the data collection process, this is an unusually successful modelling episode.

References

Aigner, D., Goldberger, A. (1977). *Latent Variables in Socio-economic Models*. North-Holland Publishing Co., Amsterdam.

Anderson, J., and Gerbing, D., (1982). "Some Methods for Respecifying Measurement Models to Obtain Unidemensional Construct Measurement," *Journal of Marketing Research*, Vol XIX, November, pp. 453-60.

Asher, H. (1976). *Causal Modeling*.. Sage Publications, Beverly Hills, CA.

Bagozzi, Richard P. (1980). *Causal Models in Marketing*, John Wiley and Sons, New York.

Bentler, P. (1990). Model Search, TETRAD II, and EQS, Sociological Methods & Research, Vol. 19, N. 1, August, pp. 67-79.

Bentler, P. (1986). *Lagrange Multiplier and Wald Tests for EQS and EQS/PC*, BMDP Statistical Software Inc., Los Angeles, California.

Bentler, P. (1985).*Theory and Implementation of EQS: A Structural Equations Program* BMDP Statistical Software Inc., Los Angeles, California.

Bentler, P. (1980). Multivariate Analysis with Latent Variables:Causal Modeling. *Annual Review of Psychology* 31:419-56.

Blalock, H. M. (1961). *Causal Inferences in Nonexperimental Research*. The University of Norh Carolina Press, Chapel Hill, North Carolina.

Blalock, H. M. (ed.) (1971). *Causal models in the Social Sciences*. Aldine-Atheston: Chicago.

Bollen, K. (1989). *Structural Equations with Latent Variables*, Wiley.

Bollen, K. (1990). "Outlier Screening and a Distribution-Free Test for Vanishing Tetrads," *Sociological Methods and Research*, Vol. 19, N. 1, August, pp. 80-92.

Callahan, J., and Sorensen, S., (1992) "Using TETRAD II as an Automated Exploratory Tool," *Sociological Methods and Research.*, Fall.

Costner, H. (1971). Theory, Deduction and Rules of Correspondence. In Blalock, H. (editor), *Causal Models in the Social Sciences.* Aldine, Chicago.

Costner, H., and Schoenberg, R. (1973). "Diagnosing Indicator Ills in Multiple Indicator Models" in Goldberger, A., Duncan, O. (editors *Structural Equation Models in the Social Sciences*, Seminar Press, New York.

Costner, H. and Herting, J. (1985). Respecification in Multiple Indicator Models. In Blalock, H. (editor), *Causal Models in the Social Sciences:2nd edition*, pp. 321-393. Aldine Publishing Co., New York.

Duncan, O. (1975). *Introduction to Structural Equation Models.* Academic Press,New York.

Fox, J. (1984). *Linear Statistical Models and Related Methods*, John Wiley and Sons, New York. 1984

Glymour, C., Scheines, R., Spirtes, P., and Kelly, K. (1987). *Discovering Causal Structure*, Academic Press, San Diego, California,.

Glymour, C., Scheines, R., Spirtes, P. (1989). "Why Aviators Leave the Navy: Applications of Artificial Intelligence Procedures in Manpower Research," *Report to the Naval Personnel Research Development Center*, January.

Glymour, C., Spirtes, P. and Scheines, R. (1991). "Causal Inference," *Erkenntnis*, Vol. 35, pp. 151-189.

Goldberg, A., Duncan, O. (editors). (1973). *Structural Equation Models in the Social Sciences.* Seminar Press, New York.

Goodman, L. A. (1973a). Causal analysis of data from panel studies and other kinds of surveys. *Amer. J. Sociol.* 78, pp. 1135-1191.

Harary, F., Norman R., Cartwright, D. (1965). Structural Models: *An Introduction to the Theory of Directed Graphs.* Wiley, New York.

Harary, F., and Palmer, E. (1973). *Graphical Enumeration*, Academic Press, New York.

Harman, H. (1976). *Modern Factor Analysis* (3rd. ed.), University of Chicago Press.

Heise, D. (1975). *Causal Analysis*, Wiley, New York, New York.

Herting, J., and Costner, J. (1985). "Respecification in Multiple Indicator Models", *Causal Models in the Social Sciences*, 2nd Edition, ed. by H. Blalock.

James, L., Mulaik, S., and Brett, J. (1982). *Causal Analysis: Assumptions, Models and Data.* Sage Publications, Beverely Hills, California.

Joreskog, K. (1973). A General Method for Estimating a Linear Structural Equation. In Goldberger, A., Duncan, O. (editors), *Structural Equation Models in the Social Sciences*. Seminar Press, New York.

Joreskog, K. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika* 43:443-447.

Joreskog, K. (1981). "Analysis of covariance structures." *Scandanavian Journal of Statististics*, Vol. 8, pp. 65-92.

Joreskog, K. and Sorbom, D. (1984). *LISREL VI User's Guide*, Scientific Software, Inc., Mooresville, Indiana.

Joreskog, K. and Sorbom, D. (1990). "Comments on: Simulation Studies of the Reliability of Computer Aided Specification Using the TETRAD II, EQS, and LISREL Programs", *Sociological Methods and Research*, August.

Kaplan, D., (1988). "The Impact of Specification Error on the Estimation, Testing, and Improvement of Structural Equation Models," *Multivariate Behavioral Research*, Vol. 23, pp. 69-86.

Kaplan, D., (1989). "A Study of the Sampling Variability and z-Values of Parameter Estimates From Misspecified Structural Equation Models," *Multivariate Behavioral Research*, Vol. 24, N. 1, pp. 41-57.

Kaplan, D., (1989a). "Model Modification in Covariance Structure Analysis: Application of the Expected Parameter Change Statistic," *Multivariate Behavioral Research*, Vol. 24, N. 3, pp. 285-305.

Kaplan, D., (1990). "Evaluating and Modifying Covariance Structure Models: A Review and Recommendation," *Multivariate Behavioral Research*, Vol. 25, N. 2, pp. 137-55.

Kenny, D. (1979). *Correlations and Causality*. Wiley, New York.

Kiiveri, H., Speed, T. (1982). Structural analysis of multivariate data: a review. In Leinhardt, S. (ed.) *Sociological Methodology*. Hoeesy Bass: San Francisco.

Kohn, M. (1969). *Class and Conformity*. Dorsey Press, Homewood, Il.

Lawley, D.N., and Maxwell, A.E. (1971). *Factor Analysis as a Statistical Method*, Butterworth, London.

Lazarsfeld, P. F., and Henry, N. W. (1968). *Latent Structure Analysis*, Houghton Mifflin, Boston.

Lee, S. (1985). Analysis of Covariance and Correlation Structures. *Computational Statistics and Data Analysis* 2:279-295.

Lee, S. (1987). *Model Equivalence in Covariance Structure Modeling*, Ph.D Thesis, Department of Psychology, Ohio State University.

Lohmoller, Jan-Bernd, (1989). *Latent Variable Path Modeling with Partial Least Squares*, Springer-Verlag.

Luijben, T., and Boomsma, A., (1988) "Statistical Guidance for Model Modification in Covariance Structure Analysis," in the Proceedings of Compstat 1988, Meetings of the International Association for Statistical Computing.

MacCallum, R. (1986). Specification Searches in Covariance Structure Modeling. *Psychological Bulletin* 100:107-120.

McDonald, R.P. (1978). "A Simple Comprehensive Model for the Analysis of Covariance Structures," *British Journal of Mathematical and Statistical Psychology*, Vol 33, pp. 161-83.

Pearl. J. (1988). *Probabilistic Reasoning in Intelligent Systems*, Morgan and Kaufman: San Mateo.

Saris, W., Satorra, A., and Sorbom, D. (1987). "The Detection and Correction of Specification Errors in Structural Equation Models," in CC. Clogg (ed.), *Sociological Methodology 1987*, San Francisco: Jossey-Bass, pp. 105-129.

Saris, W., Pijper, W., and Zeqwaart, P., (1979). "Detection of Specification Errors in Linear Structural Equation Models," in *Sociological Methodology 1979*, K. F. Schuessler, ed. SanFrancisco: Jossey-Bass Inc., Publishers

Scheines, R., Spirtes, P., Glymour, G., and Sorensen, S. (1990). "Causes of Success and Satisfaction Among Naval Recruiters," *Report to the Naval Personnel Research Development Center*, July.

Scheines, R., Spirtes, P., and Glymour, C. (1990a). "A Qualitative Approach to Causal Modeling", in P. Fishwick, ed., *Qualitative Simulation Modeling*.

Scheines, R., Spirtes, P., and Glymour, C. (1991), "Building Latent Variable Models," LCL-Technical Report 91-2, Carnegie Mellon University, Pittsburgh, PA.

Scheines, R., and Spirtes, P. (1992). "Finding Latent Variable Models in Large Data Bases", forthcoming in G. Piatetski-Shapiro, ed., *International Journal of Intelligent Systems*.

Sorbom., D. (1975). "Detection of Correlated Errors in Longitudinal Data," *British Journal of Mathematical and Statistical Psychology* 28:138-151.

Sorbom., D. (1989). "Model Modification," *Psychometrika*, V. 54, pp. 371-84.

Spirtes, P. (1989). "A Necessary and Sufficient Condition for Conditional Independencies to Imply a Vanishing Tetrad Difference." Technical Report CMU-LCL-89-3, Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P. (1989a). "Fast Geometrical Calculations of Overidentifying Constraints", Carnegie-Mellon University Laboratory for Computational Linguistics Technical Report CMU-LCL-89-3.

Spirtes, P., and Glymour, C. (1988). "Latent Variables, Causal Models and Overidentifying Constraints", *Journal of Econometrics*, 39, pp. 175-198, Sept..

Spirtes, P. and Glymour, C. (1990). "An Algorithm for Fast Recovery of Sparse Causal Graphs." Technical Report CMU-LCL-90-5, Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P. and Glymour, C. (1990a). "Causal Structure Among Measured Variables Preserved with Unmeasured Variables." Technical Report CMU-LCL-90-5, Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P., Glymour, C., and Scheines, R. (1990c). "Causality from Proability", in G. McGee, ed., *Evolving Knowledge*, Pitman.

Spirtes, P., Glymour, C., and Scheines, R., and Sorensen, S. (1990e). TETRAD Studies of Data for Naval Air Traffic Controller Trainees, *Report to the Naval Personnel Research Development Center*, January.

Spirtes, P., Scheines, R., and Glymour, C. (1990f). "Simulation Studies of the Reliability of Computer Aided Specification Using the TETRAD II, EQS, and LISREL Programs", *Sociological Methods and Research*, pp. 3-60.

Spirtes, P., Glymour, C., and Scheines, R (1991). "An Algorithm for Fast Recovery of Sparse Causal Graphs." *Social Science Computer Review.*

Spirtes, P., Scheines, R., Glymour, C., and Meek, C. (1992a). "TETRAD II User's Manual", Unpublished Manuscript, available from the Dept. of Philosophy, Carnegie Mellon University.

Spirtes, P., Scheines, R., and Glymour, C. (1992). *Causation, Prediction, and Search*, forthcoming in the series: Lecture Notes in Statistics. Springer-Verlag.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, Wiley: New York.

Wishart, J. (1928). "Sampling Errors in the Theory of Two Factors," *British Journal of Psychology*, Vol. 19, pp. 180-187.

Wright, S. (1934). "The Method of Path Coefficients", *Annals of Mathematical Statistics*, 5, pp. 161-215.

Appendix A. The MIMBuild Algorithm

A1. The Algorithm

MIMBuild first purifies the measurement model, and then computes a class of structural models that strongly imply the vanishing correlations and vanishing first order partial correlations among the latent variables. Its output for the second stage is a *modified pattern* Π, i.e., partially directed acyclic graph in which several sorts of adjacencies can occur:

A--B
A->B
A?-B
A?>B

The algorithm's strategy is similar to the PC algorithm in the Build module. It begins with a complete undirected graph, prunes adjacencies, and then orders those that remain.

<u>The MIMBuild Algorithm</u>

Begin: For every pair of latent variables X,Y, let X--Y be in Π.
Eliminate an edge X--Y from Π just in case either
i) $\rho_{XY} = 0$, or
ii) there exists a Z distinct from X and Y such that $\rho_{XY.Z} = 0$.

<u>Identifying Causal Order</u>

Begin: Let Π be the output from Forming the Adjacencies.
1) Forming Colliders
For every triple X, Y, Z such that exactly X--Y--Z is in Π, i.e.,
i) X--Y is in Π,
ii) Y--Z is in Π, and
iii) X--Z is not in Π,
if a) $\rho_{XZ} = 0$, or
b) $\rho_{XZ} \neq 0$ and $\rho_{XZ.Y} \neq 0$,
then order X--Y--Z as X->Y<-Z

2) Avoiding Colliders

Repeat

For every triple X, Y, Z such that exactly X--Y<-Z is in Π,

if $\rho_{XZ.Y} = 0$, then X--Y is X<-Y.

Until an iteration is completed in which nothing changes

Marking Adjacencies

If X and Y are adjacent in Π, and there is some path of length ≥ 2 connecting X and Y in Π, and there exists no vertex Z such that

a) Z occurs on all undirected paths **P** of length ≥ 2 connecting X and Y in Π, and

b) there do not exist two undirected paths P_i and $P_j \in$ **P** such that Z could occur as a collider on P_i and a non-collider on P_j

then convert X--Y to X?-Y, or convert X->Y to X?>Y, or convert X<-Y to X<?Y.

A2. Theoretical Reliability and Complexity

Supposing that G is a pure latent variable model parameterized such that every vanishing correlation and vanishing tetrad difference is strongly implied to vanish, then the following theorem about MIMBuild's theoretical reliability is proved in (Spirtes 92).

Theorem 1: If G is a pure latent variable model in which each latent variable has at least two measured indicators, and MIMBuild is given the set of vanishing correlations and vanishing tetrad differences strongly implied by G, then its output Π is correct in the following respects about G.

Adjacency

A-1) If X and Y are not adjacent in Π, then they are not adjacent in G.

A-2) If X--Y or X->Y is in Π, then X and Y are adjacent in G.

Ordering

O-1) If X?>Y is in Π, then every trek in G between X and Y is into Y.

O-2) If X->Y is in Π, then X->Y is in G.

The algorithm's complexity is bounded by the number of tetrad differences it must check, which in turn is bounded by the number of foursomes of measured variables. If there are n measured variables the total number of foursomes is $O(n^4)$. We don't check each possible foursome, however, and the actual complexity depends on the number of latent variables and how many variables measure each latent. If there are m latent variables and s measured variables for each, then the number of foursomes is $O(m * s^4)$.[10] Since $m*s = n$, this is $O(n * s^3)$.

[10] See (Scheines 91).

Appendix B. The YATS Model

Using TETRAD II as an Automated Exploratory Tool

by

Jan Callahan
Callahan Associates

Stephen W. Sorensen
Navy Personnel Research and Development Center

to appear in:

Sociological Methods and Research
Fall 1992

Received:

Certificate of Commendation
from
The Chief of Naval Research
in recognition of nomination for
Best FY-90 Independent Exploratory Development Project

USING TETRAD II AS AN AUTOMATED EXPLORATORY TOOL

JANICE D. CALLAHAN

Callahan Associates Incorporated

STEPHEN W. SORENSEN

Navy Personnel Research and Development Center

ABSTRACT

An automated process for forming measurement models and structural equation models was developed using an experimental version of TETRAD II. An evaluation function was developed to select the measurement models and the structural models for further consideration. Comparisons were made between developing the models in this way and using factor analysis to develop measurement models. We found TETRAD II useful in building sets of measurement models and in suggesting sets of possible structural models. A previously unstudied dataset with a sample size of 7625 was used for testing the process. We developed measurement models and structural equation models which satisfied our intuitive understanding and had a Chi-square goodness-of-fit p-value of 0.77.

Keywords: EQS, structural models, measurement models, model specification, TETRAD.

INTRODUCTION

Researchers at the Navy Personnel Research and Development Center (NPRDC) are often requested to analyze large datasets that they did not design or collect. Occasionally, questions must be answered for which the data were not designed. As a result, we have been studying techniques for discovering information from large databases about which little is known. As part of this project, we have been using and evaluating the TETRAD II program (Spirtes et al., 1992).

TETRAD was designed as an aid in elaborating or respecifying structural equation models (Glymour et al., 1987). Spirtes, Scheines and Glymour (1990) have shown that TETRAD II's elaborator performs well in simulations. Given a correlation matrix and a structural equation model, the elaborator then finds alternate sets of structural equation models that satisfy the correlation constraints of the data and the model.

We used two experimental modules of TETRAD II to form measurement and structural equation models. We automated the selection and evaluation of the models. The result of this process is a set of tentative models which can be used as a basis for further research.

SPECIFYING A MODEL

Only recently have statisticians begun to address model specification (Lehmann, 1990). Stepwise regression and all possible models methods (Draper & Smith, 1981) are data-driven techniques developed to reduce dimensionality and investigate the "best" model. Edwards & Havránek (1985) extended all possible models techniques to quickly find multiple regression models for large, complex problems. A large literature has developed on graphical techniques for dimension reduction and model building (Chambers, Cleveland, Kleiner and Tukey, 1983). Cook and Weisberg (1990) have developed interactive graphical techniques for model selection that include not only variable selection but also the functional form of those variables. Weihs and Schmidli (1990) have developed an interactive graphical tool for multivariate exploratory analysis.

CART and CHAID are non-parametric computerized techniques which develop tree models for predicting a categorical outcome variable. Edwards and Havránek (1987) presented techniques for searching for models in multidimensional contingency tables.

In the spirit of these model specifying techniques we decided to use TETRAD II to generate sets of measurement and structural equation models for latent variables. To the best of our knowledge, this is the first technique available for locating sets of possible structural equation models. Substantive knowledge complements this technique in three ways: selecting variables to measure, choosing variables as possible indicators of a latent variable, and rejecting nonsensical models.

METHODS

We used a large dataset with which we were unfamiliar. We developed measurement models for latent variables and then built structural models detecting causal relations among such latent variables. Measurement models were formed in two ways: clustering of variables and factor analysis. TETRAD II was used to select subsets that form pure measurement models. After selecting measurement models, structural models were built using TETRAD II. We evaluated all the measurement models and structural models using the EQS program.

Measurement Models Formation - Theory

An experimental module of TETRAD II, SCALES, was used to form pure measurement models. A pure measurement model is one in which all correlations between the indicator variables, $\varrho_{Xi,Xj}$, are due solely to the common effects of a single latent variable. Given a group of possible indicators of one latent variable, SCALES searches for five-variable subsets that are pure measurement models for that latent variable. High correlations are expected among the variables in this group; variables with zero correlations are eliminated.

A pure measurement model implies that all the tetrad equations are true. That is, for all sets of four variables in the measurement model, X_i, X_j, X_m, and X_n, the tetrad equation, $\varrho_{i,j} * \varrho_{m,n} - \varrho_{i,m} * \varrho_{j,n} = 0$, is true. $\varrho_{i,j}$ is the true correlation between Xi and Xj in the population. If there are five variables in the measurement model, then there are 15 tetrad equations.

SCALES judges the pureness of a measurement model by evaluating the 15 sample tetrad equations from each fivesome (Glymour and Spirtes, 1988). The residual from each sample tetrad equation, $r_{i,j} * r_{m,n} - r_{i,m} * r_{j,n}$, is calculated, where $r_{i,j}$ are sample correlations. On the assumption that each equation holds in the population, SCALES calculates the probability that the residual from the tetrad equation is as large as or larger than, the one observed. This is equivalent to a p-value for the hypothesis that the residual equals zero. We would like these probabilities to be large. SCALES prints out the minimum and the average of the 15 probabilities.

SCALES also prints out the proportion of tetrad equations that can only be explained by a latent variable. If there exists a measured variable, X_p, such that all partial correlations partialled on X_p ($\varrho_{i,j,p}$ for i, j=1, 2, 3, 4) are zero, then the tetrad equation among X_1, X_2, X_3, X_4 is true. This means a tetrad equation can only be explained by a latent variable if there does not exist another measured variable, X_p, such that all partial correlations partialled on X_p are zero (Spirtes, 1989).

SCALES calculates the above three assessments for every possible fivesome and foursome. For a more detailed discussion of the SCALES module, see Scheines, et al., 1991.

Measurement Models Formation - Procedure

Variables were clustered as indicators of some latent variable by two methods: grouping questions that addressed the same issue and factor analysis (Bollen, 1989). We chose PROMAX rotations because these rotations made the most sense. The factor analyses often found only 2 or 3 variables for each factor. These factors could not be evaluated by SCALES, because SCALES requires at least 5 variables. Nor could these factors be evaluated by EQS, because with only 2 or 3 variables the models are underidentified. A third technique is available for evaluating measurement models: a TETRAD II module called PARTIAL. However, we exceeded the number of variables that PARTIAL could analyze. As a result, those factors had to be dropped from further analysis. All factor analyses were performed using SAS on an IBM4381.

After giving SCALES a cluster of variables and receiving sets of foursomes and fivesomes, we ranked these sets according to the sum of the 3 assessments discussed above. The top 20 sets (many latent variables did not have that many) were then evaluated by EQS using SCALES-generated EQS code. For each latent variable, we selected as the measurement model the foursome or fivesome with the highest p-value for the Chi-square goodness-of-fit test from EQS. We were careful to insure that each variable was included in only one measurement model. The same variable in two measurement models would violate a pure measurement model.

Structural Equation Models Development

We estimated the correlations among the latent variables with EQS. Path models were then constructed with another experimental module of TETRAD II, PARTIAL. The PARTIAL module of TETRAD II eliminates causal connections between two variables by using partial correlations as statistical tests for conditional independence. If two variables are independent conditional on some other set of variables, then the two variables are not directly causally connected (Spirtes et al., 1991). Thus, through a process of elimination, PARTIAL produces a set of possible structural models. We evaluated all these structural models for goodness of fit with EQS.

The Dataset Used

The data we used consisted of responses to the Youth Attitude Tracking Survey (YATS). Each year this questionnaire is given to about 10,000 young persons between the ages of 18 and 25. These individuals are tracked to see if they enlist in the military. The purposes of the questionnaire are to evaluate the effectiveness of recruitment advertising and to look for any enlistee-identifying characteristics. We used the responses from the 1985 questionnaire because five years had elapsed during which respondents could have enlisted.

Many of the questions in YATS can be answered yes or no. We derived variables that approximate continuous measures for use in TETRAD II. We were primarily interested in identifying any differences between those who enlisted and those who did not. Rather than include this binary variable in any model, we divided the responses into two datasets, military and civilian and analyzed each separately. There were 7625 civilian respondents (i.e., respondents with no enlistment to date) and 854 individuals who had enlisted.

RESULTS

Table 1 contains an example of a common-sense grouping of the questions that we believed addressed "how likely a respondent felt he or she was to enlist in some branch of the military in the future" (the Likely Military latent variable). Table 2 contains the best 3 measurement models of fivesomes and the best 4 models of foursomes from this set, where "best" means the highest values for the sum of the 3 TETRAD II probabilities. The second through fourth columns contain the TETRAD II assessment probabilities. The rightmost column contains the p-level from the goodness-of-fit Chi-square test from EQS. Notice that the goodness-of-fit statistics agree in general with the TETRAD II statistics. In almost all cases, the model with the highest sum of the three TETRAD II assessments also had the highest p-value from EQS. Occasionally the model with the highest p-value from EQS had the second highest sum of the three TETRAD II assessments.

The PARTIAL module provided 3 to 20 structural models for each dataset (civilian, military, clusters, factor analysis). Most of these models differed only in the causal direction between one or more pairs of latent variables. Some latent variables were never causally connected to others. These latent variables were dropped from the structural model. Usisng Chi-square goodness-of-fit p-values from EQS to evaluate each of the suggested structural models, none of the structural models using factor analysis measurement models fit well: The highest p-value for the Chi-square goodness of fit was .0201 and most p-values were <.001. Similarly, the structural models for military enlistees using substantive groupings for measurement models all had p-values <.001.

However, for the civilians using substantive groupings for the measurement models, the process found consistent and revealing models. Starting with seven latent variables, TETRAD II suggested 20 possible structural models among five latent variables, dropping two. Some definite patterns emerged from studying the 20 models. Figure 1 summarizes the 20 path models. The latent variables have been abbreviated as follows: CURrent JOB, FRIENDS support military, FUTure JOB, FUTure MILitary plans, LIKELY military. The heavy-lined edges are required for a good fit. Any path models without these edges, or with any of these edges pointing in the opposite direction, had very low p-values. The light-lined edges were included in models that fit well, but the direction of these edges was unclear. However, no more than one of the light-lined edges could have the direction switched from Figure 1 and still fit well.

Specifically, four structural models had Chi-square goodness-of-fit p-values of 0.77. All four models contained the five heavy-lined edges with the coefficients and directions indicated. One model had all the edges and directions displayed. The other three models had one of the three light edges pointing in the other direction. Two more models had two of the three light edges pointing in the opposite directions; they had p-values near 0.4. All other path models had very low p-values, $<=.05$.

Many of the heavy-lined edges make sense. Individuals' current jobs impact on their thoughts about their future jobs. And how the individuals feel about possible future non-military jobs is causally connected to how they feel about a future in the military. The negative coefficient for that edge means that the more positive individuals are about a future non-military job the less they see a future in the military. Similarly, if friends or relatives are positive about the military, or if the friends or relatives think positively about a future military career, then the respondents are more likely to see a future for themselves in the military.

CONCLUSIONS

We found TETRAD II a useful method for uncovering and evaluating tentative measurement models and for building and eliminating structural equation models on the latent variables. We began with a new database and were able to generate structural equation models that fit remarkably well. TETRAD II is very helpful in indicating what does not work. Variables that never appear in a foursome or fivesome are not useful as an indicator for that latent variable. Also, if a latent variable does not appear in any suggested structural equation model, that latent variable is not useful.

We were impressed with the quality of the model that TETRAD II produced. Could a human researcher have found a model as good as that found by the computer? We don't know. Although we had to write computer programs to link TETRAD II and EQS, the computer made its own decisions as it searched for the model. The computer acted solely from the evaluation functions that we provided.

Statisticians and most researchers are aware that if you look hard enough you will find a model that fits a dataset. This model, however, may not fit a new dataset well. Due to the optimizing, stepwise regression and all-models methods almost always understate errors for future datasets. CART and CHAID have techniques for correcting for the optimization. Our technique with TETRAD II has no built-in technique for adjusting p-levels for the optimizing. Test samples, cross validation or bootstrapping techniques (Efron and Tibshirani, 1991) can be used with our TETRAD II technique to correct for optimizing.

REFERENCES

Bollen, K. 1989. Structural equations with latent variables. New York, NY: Wiley.

Chambers, J., Cleveland, W., Kleiner, B. & Tukey, P. 1983. Graphical methods for data analysis. Belmont, CA: Wadsworth.

Cook, R. & Weisberg, S. 1990. Added variable plots in linear regression. In Proceedings of IMA Conference on Robustness and Diagnostics. New York, NY: Springer-Verlag.

Draper, N. & Smith, H. 1981. Applied regression analysis, second edition. New York, NY: Wiley.

Edwards, D. & Havrânek, T. 1985. A fast procedure for model search in multidimensional continengcy tabies. Biometrika, 72: 339-351.

Edwards, D. & Havrânek, T. 1987. A fast model selection procedure for large families of models. Journal of the American Statistical Association, 82: 205-213.

Efron, B. & Tibshirani, R. 1991. Statistical data analysis in the computer age. Science, 253: 390-395.

Glymour, C., Scheines, R., Spirtes, P. & Kelly, K. 1987. Discovering Causal Structure. San Diego, CA: Academic Press.

Glymour, C. & Spirtes, P. 1988. Latent Variables, Causal Models and Overidentifying Constraints. Journal of Econometrics, 39: 175-198.

Lehmann, E. 1990. Model specification: the views of Fisher and Neyman, and later developments. Statistical Science, 5: 160-168.

Scheines, R., Spirtes, P. Glymour, C. 1991. Building Latent Variable Models. Technical Report LCL-91-2. Carnegie Mellon University.

Spirtes, P. 1989. A Necessary and Sufficient Condition for Conditional Independencies to Imply a Vanishing Tetrad Difference. Technical Report CMU-LCL-89-3, Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P., Glymour, C., Scheines, R. 1991. An Algorithm for Fast Recovery of Sparse Causal Graphs. Social Science Computer Review, in press.

Spirtes, P., Scheines, R., Glymour, C. 1990. Simulation Studies of the Reliability of Computer-Aided Model Specification Using the TETRAD II, EQS, and LISREL Programs, Sociological Methods and Research, 19: 3-67.

Spirtes, P., Scheines, R., Glymour, C. 1992. Causality, Prediction and Search. New York, NY: Springer-Verlag, Lecture Notes in Statistics, in press.

Weihs, C. & Schmidli, H. 1990. OMEGA Online multivariate exploratory graphical analysis: routine searching for structure. Statistical Science, 5: 175-226.

SOFTWARE CITED

CART. California Statistical Software, Inc., 961 Yorkshire Ct., Lafayette, CA 94549; 415-283-3392.

EQS. Available from BMDP Statistical Software, Inc., 1440 Sepulveda Blvd., #316, Los Angeles, CA 90025; 213-479-7799.

SAS. SAS Institute, POB 8000, Cary, NC 27512; 919-467-8000.

SPSS/PC+ CHAID. SPSS, Inc. 444 N. Michigan Avenue, Chicago, IL 60611; 312-329-3300.

ENDNOTE: This research was partly performed under contract N66001-88-D-0054 that Systems Engineering Associates, San Diego, CA, has with the Navy Personnel Research and Development Center. The opinions expressed in this paper are those of the authors, are not official, and do not necessarily reflect the views of the Navy Department.

ACKNOWLEDGEMENTS: We thank Peter Spirtes, Clark Glymour, and especially Richard Scheines for their advice, suggestions, and patience. We also thank the reviewers for their helpful comments and suggestions.

BIOGRAPHICAL SKETCHES: Janice D. Callahan is a statistical consultant with Callahan Associates Incorporated. Her interests are in biotechnology, marine ecology, and statistical analysis of large data bases. Stephen W. Sorensen works for the Navy Personnel Research and Development Center in the planning of Navy training. His research interest is the application of artificial intelligence and operations research to decision support systems.

Table 1. Cluster of Questions for the "Likely Military" Latent Variable

Variable	Question
Q505	How likely is it that you will be serving in the National Guard?
Q507	How likely is it that you will be serving in the Reserves?
Q509	How likely is it that you will be serving on active duty in the Coast Guard?
Q510	How likely is it that you will be serving on active duty in the Army?
Q511	How likely is it that you will be serving on active duty in the Air Force?
Q512	How likely is it that you will be serving on active duty in the Marine Corps?
Q513	How likely is it that you will be serving on active duty in the Navy?
Q514	How likely is it that you will be going to college?
Q515	How likely is it that you will be going to vocational or technical school?

Table 2. SCALES Selections for "Likely Military" Latent Variable for Civilian Respondents

Variables	Average Prob.	Minimum Prob.	Prob. Needs Latent	EQS P(Chi Sq)
Q507 Q509 Q510 Q511 Q515	0.633	0.000	0.87	0.0048
Q507 Q509 Q511 Q512 Q515	0.340	0.011	1.00	0.0305
Q507 Q510 Q511 Q513 Q515	0.377	0.000	0.80	<.001
Q507 Q509 Q511 Q515	0.835	0.750	1.00	0.9485
Q507 Q510 Q511 Q515	0.704	0.570	1.00	0.8443
Q509 Q510 Q511 Q515	0.688	0.532	1.00	0.8208
Q507 Q509 Q510 Q515	0.679	0.537	1.00	0.8469

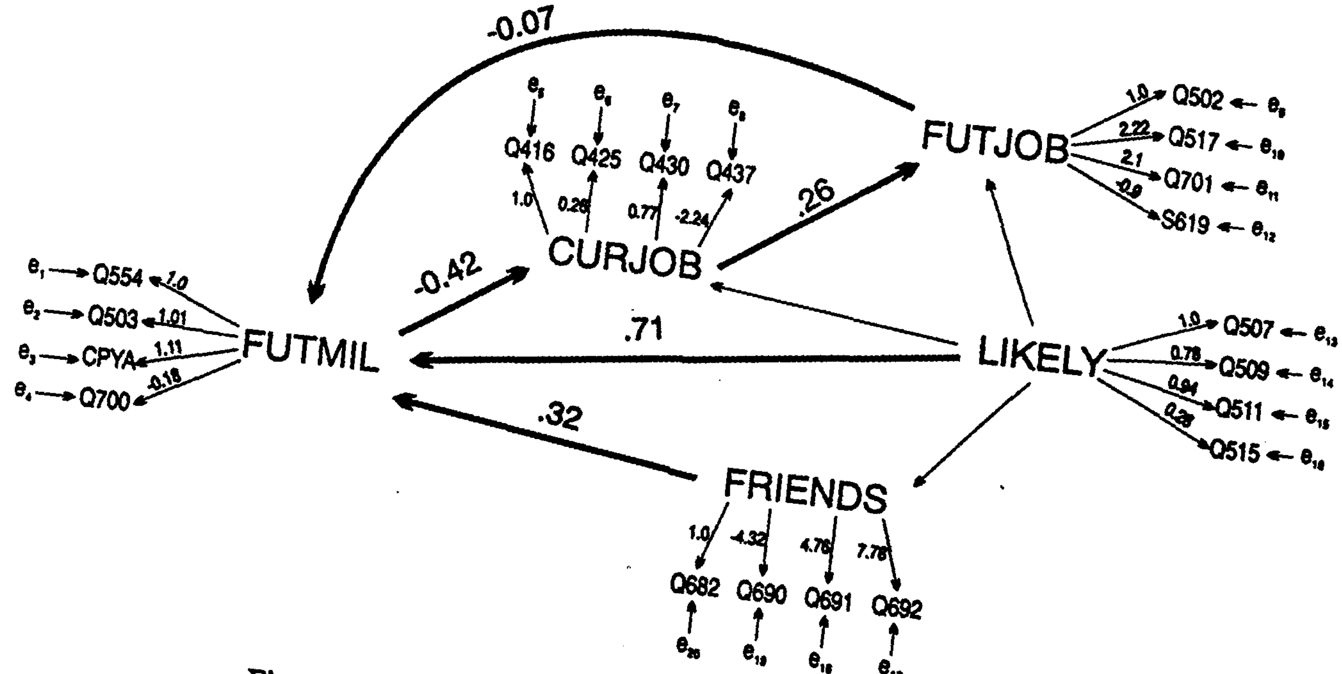

Figure 1. Civilian Structural Model, n = 7625, p = .77